International Journal of Biomedicine 8(4) (2018) 284-287
http://dx.doi.org/10.21103/Article8(4)_OA2

ORIGINAL ARTICLE

Cardiology

INTERNATIONAL JOURNAL OF BIOMEDICINE

A Multiple Logistic Regression Model as an Additional Mathematical Method for Predicting the Development of Ischemic Stroke in Patients with Atrial Fibrillation

Natalya V. Aksyutina, PhD, MD[1*]; Vladimir A. Shulman, PhD, ScD[1]; Svetlana Yu. Nikulina, PhD, ScD[1]; Vasily V. Kozlov, PhD, MD[2]; Oksana A. Gavrilyuk, PhD[1]; Andrei V. Bespalov, MD[1]; Larissa G. Soinova[1]; Eduard V. Semyonov, PhD, MD[1]

[1]*Professor V.F. Voino-Yasenetsky Krasnoyarsk State Medical University, Krasnoyarsk, the Russian Federation*
[2]*I.M. Sechenov First Moscow State Medical University, Moscow, the Russian Federation*

Abstract

Prevention of thromboembolic complications in cases of atrial fibrillation (AF) and, above all, ischemic stroke (IS), represents a key problem of modern cardiology. ***The aim*** of the present study was to assess the feasibility of Multiple Logistic Regression Analysis in predicting the occurrence of IS in AF patients with the predictor genotypes of the *FGB*, *GPIα*, and *GPIβα* genes, in order to implement an approach to primary prevention and personalized treatment.

Methods and Results: We examined 43 patients with atrial fibrillation and IS in their histories and 78 patients with AF without IS. A total of 188 persons without AF were included in the control group. The present study showed that the homozygote minor allele genotype (AA) of the *FGB* -455G/A SNP, the minor allele CT and TT genotypes of the *GPIa* 807C/T SNP, and the -5C/-5C and -5C/-5T genotypes of the *GPIβα* -5T/C polymorphism can be studied as genetic predictors of IS in AF patients. Logistic regression analysis was used to predict the development of IS in AF patients, depending on the presence of pathological genotypes of the *FGB*, *GPIα*, and *GPIβα* genes. The percentage of correct predictions for the absence of IS using this model was 99.5%. The development of IS was correctly predicted in 7.0% of cases. The overall percentage of correct predictions was 82.3%.

Conclusion: The obtained logistic regression model is recommended as an additional method for assessing the risk of IS in young patients with lone AF. (**International Journal of Biomedicine. 2018;8(4):284-287.**)

Key Words: atrial fibrillation • ischemic stroke • multiple logistic regression analysis • predictor genotypes

Abbreviations

AF, atrial fibrillation; **ACA**, acute cerebrovascular accident; **AUC**, area under the curve; **FGB**, fibrinogen beta chain; **IS**, ischemic stroke; **MLRA**, Multiple Logistic Regression Analysis; **OR**, odds ratio; **CI**, confidence interval. **SNP**, single nucleotide polymorphism; **GPIa**, glycoprotein Ia; **GP1βα**, glycoprotein Ib platelet subunit alpha.

Introduction

Prevention of thromboembolic complications in cases of AF and, above all, IS, represents a key problem of modern cardiology.[1-7] The formation of AF-related blood clots in the atria is a result of a complex interaction among different factors, including the enlargement of the left atrial appendage, hemostasis, endothelial dysfunction, systemic and, conceivably, local hypercoagulation.[8] The CHA2DS2-VASc score for Stroke Risk Assessment in AF presents quantitative estimates of the various clinical risk factors.[9,10] However, over the past decade a number of studies have discussed the genetic risk factors for IS development in patients with AF.[11-15] Mainly, the objects of the study were the genes of the

***Corresponding author:** Natalya V. Aksyutina, PhD, MD. Krasnoyarsk State Medical University n.a. Prof. V.F. Voino-Yasenetsky, Krasnoyarsk, the Russian Federation. E-mail: aks-n-v@yandex.ru*

hemostasis system, but more generally, the objects of study were the genes of the hemostatic system.

Several polymorphisms of platelet membrane glycoprotein have been identified as potential risk factors for cardiovascular disease. Thus, a nucleotide −5T/C dimorphism in the translation initiation site (Kozak sequence) of the GPIbα gene has been associated with increased platelet surface levels of the GPIb-IX-V receptor complex.[16] The role of this GPIbα Kozak sequence polymorphism in the occurrence of arterial thrombotic disease is known.

Platelet *GPIa* 807C/T is the only GP polymorphism associated with the expression levels of GP Ia/IIa (the platelet collagen receptor). Logistic regression analysis revealed that the presence of the *GPIa* 807C/T C allele and CC genotype were both associated with a decreased risk of CH compared with T allele, CT and TT genotypes.[17]

Xiaofeng Hu et al. found that the *FGB* -455 G/A polymorphism was independently associated with increased risk of cardioembolic stroke in AF patients with a low CHA_2DS_2-VASc score.[18]

The researchers presented data on the association of predictor genotypes with the development of IS in AF patients.[19] At the same time, the use of mathematical methods for predicting this complication in AF patients is relevant for the isolation of high-risk groups among patients and the implementation of measures for targeted prevention and use of personalized treatment.

The aim of the present study was to assess the feasibility of MLRA in predicting the occurrence of IS in AF patients with the predictor genotypes of the *FGB*, *GPIα*, and *GPIβα* genes, in order to implement an approach to primary prevention and personalized treatment.

Materials and Methods

We examined 43 patients with AF and IS in their histories (Group 1) and 78 patients with AF without IS (Group 2). A total of 188 persons without AF were included in the control group (Group 3). The median age in the groups did not differ significantly (58 years [52; 65], 62 years [44.75; 71], and 59 years [53; 65.75], respectively). The diagnosis of AF was based on electrocardiograms (ECG) and/or Holter ECG data following standard diagnostic criteria. All patients underwent the following examinations: ECG, echocardiography, Holter ECG, exercise stress test, transesophageal stimulation of the left atrium, and blood test for thyroid hormones (TSH, free T3, free T4). To confirm the ischemic nature of acute cerebrovascular accident in the examined patients, a brain MRI was used. All participants underwent a molecular-genetic examination.

The present study was approved by the local Ethics Committee of Professor V.F. Voino-Yasenetsky Krasnoyarsk State Medical University. Written informed consent was obtained from each patient.

Statistical analysis was performed using the Statistica 7.0 software package (Stat-Soft Inc., USA) and IBM SPSS 20 (*SPSS* Inc, Chicago, IL). The normality of distribution of continuous variables was tested by the Shapiro-Wilk test. For descriptive analysis, results are presented as mean±standard deviation, median, interquartile range (IQR; 25th to 75th percentiles). Chi-squared and Fisher's exact tests were used to determine the association between categorical measure including allele and genotype. Three exact probability tests for departure from HWE due to heterozygote excess, heterozygote deficit and omnibus probability test were carried out using GENEPOP (v.4.7.0). The strength of the associations was expressed as odds ratio (OR) with 95% confidence interval (CI). Logistic regression analysis was used to predict the development of IS in AF patients, depending on the presence of pathological genotypes of the *FGB*, *GPIα*, and *GPIβα* genes. As a response, a binary variable was considered, where *0* is the absence of the predicted state (IS), *1* - its presence. The quality of the model (specificity and sensitivity) was measured by receiver operating characteristic (ROC) analysis. For all tests, a probability value of $P<0.05$ was considered statistically significant.

Results

The homozygote minor allele genotype (AA) of the *FGB* -455G/A SNP tended to elevate the risk of IS (OR=1.7, 95% CI: 1.08–2.82) in AF patients. Both minor allele TT and CT genotypes of the *GPIa* 807C/T SNP were associated with an increased risk of IS in AF patients (OR=2.5, 95% CI: 1.17-5.36). Analysis of −5T/C polymorphism of the *GPIbα* gene revealed that the rare -5C allele tended to elevate the risk of IS (OR=1.9, 95% CI: 1.05–3.42), and both -5C/-5C and -5C/-5T genotypes were associated with an increased risk of IS in AF patients (OR=2.3, 95% CI: 1.15–4.57).

Thus, the present study showed that the homozygote minor allele genotype (AA) of the *FGB* -455G/A SNP, the minor allele CT and TT genotypes of the *GPIa* 807C/T SNP, and the -5C/-5C and -5C/-5T genotypes of the *GPIβα* −5T/C polymorphism can be studied as genetic predictors of IS in AF patients. Therefore, for the further development of mathematical models to predict IS, depending on the predictor genotypes in AF patients, we used the AA genotype of the *FGB* -455G/A SNP, the TT+CT genotypes of the *GPIa* 807C/T SNP, and the -5C/-5C+-5C/-5T genotypes of the *GPIβα* gene.

The logistic regression model is a dependence of the logarithm of the chance of the predicted event (logit) on a linear combination of factor variables. Accordingly, the probability that the predicted event will occur can be represented by the following equation: $p = \frac{1}{1 + e^{-(b_0 + b_1 x_1 + \ldots + b_n x_n)}}$, where

p - the probability of the predicted event,

e - the mathematical constant 2.72 (the base of the natural logarithms),

b_0 - the constant of the model, b_1 is the coefficient of the predictor variable

x_1, - the change in logarithmic chances caused by a single change in independent variables,

n - the ordinal number of the predictor included in the equation.

When predictors were included in the multiple logistic regression equation, their collinearity and autocorrelation were tested.

The logistic regression model was built using step-by-step

inclusion of prognostic factors and determination of the minimum set of predictors by calculating Nagelkerke's R^2 to indicate the effect of all model predictors on response variable dispersion.

The statistical significance of the model was verified using the $\chi 2$ criterion. At a value of $P<0.05$, the hypothesis of insignificance of the model was rejected. Compliance of the model with the data was characterized by the Hosmer-Lemeshow goodness-of-fit test (HL test). When $P>0.05$, the hypothesis of model consistency was accepted.

The interpretation of the logistic regression parameters was based on the value of *exp(b):* if the coefficient *b* is positive, then *exp(b)* is greater than 1 and the chances of the predicted event increase; if the coefficient is negative, the chances decrease.

Sensitivity and specificity of predictors were evaluated by ROC-analysis. Quantitative interpretation of the results for the final model was carried out on the ROC-curve with the assessment of AUC.

To assess the impact of these genotypes on the probability of IS, a multiple logistic regression model was constructed:

$$p = \frac{1}{1 + e^{-(-2{,}519 + 0{,}94x_1 + 0{,}927x_2 + 0{,}854x_3)}}$$, where

p - the probability of ischemic stroke

x1- presence of the AA genotype of the *FGB* -455G/A SNP (yes – 1, no – 0), (b_1= 0.94)

x2 - presence of allele T (TT or CT genotypes) of the *GPIa* 807C/T SNP (yes – 1, no – 0), (b_2= 0.927)

x3 - presence of allele -5C (-5C/-5C or -5C/-5T genotypes) of the *GPIβα* −5T/C polymorphism (yes – 1, no-0), (b_3= 0,854),

constant b_0 = - 2,519.

The model was statistically significant. The significance of the model at the third step of the predictors' inclusion corresponded to P=0.002. The HL test demonstrates the consistency of the model with the original data (P=0.054). The data of the model is presented in Table 1.

The model reflects an increase in the possibility of IS in AF under the influence of 3 studied genetic factors. The coefficient of determination of the multiple model is R^2=0.102, which shows a statistically significant explanation of the influence of selected genetic factors on the probability of IS in the study group of patients by 10.2%. The obtained model has a high specificity. The percentage of correct predictions for the absence of IS using this model was 99.5%. The development of IS was correctly predicted in 7.0% of cases. The overall percentage of correct predictions was 82.3%.

According to the ROC curve analysis, AUC was 0.671±0.050 (95% CI: 0.574-0.768; P<0.001), which corresponds to the average quality of the model for predicting the development of IS in AF (Fig.1).

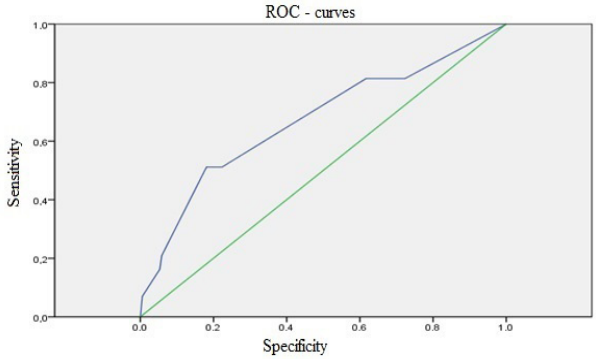

Fig. 1. *ROC - curve of the regression model of IS in AF patients depending on the presence of the predictor genotypes.*

According to the overall results of our study, the probability of IS in AF, predicted by the method of logical regression, is 7.0%; for the absence of IS - 99.5%. The total percentage of correct predictions is 82.3%.

Table 1. Linear regression model of IS depending on the presence of pathological genotypes

		Regression coefficient (b)	Standar d Error	Wald test χ^2	Sig. (*P*)	Exp (b)	95% CI for Exp(b)	
							Lower Bound	Upper Bound
Step 1	*GPIα* gene (TT and CT genotypes)	0.916	0.390	5.515	0.019	2.498	1.163	5.364
	Constant	-2.092	0.335	38.950	<0.001	0.123		
Step 2	*GPIα* gene (TT and CT genotypes)	0.906	0.394	5.293	0.021	2.474	1.144	5.350
	GPIα gene (CC and CT genotypes)	0.817	0.358	5.194	0.023	2.263	1.121	4.568
	Constant	-2.350	0.365	41.533	<0.001	0.095		
Step 3	*FGB* gene (AA genotype)	0.940	0.463	4.125	0.042	2.560	1.033	6.344
	GPIβα gene (-5T/-5T and -5C/-5T genotypes)	0.927	0.398	5.423	0.020	2.526	1.158	5.510
	GPIβα gene (-5C/-5C and -5C/-5T genotypes)	0.854	0.363	5.528	0.019	2.348	1.153	4.785
	Constant	-2.519	0.384	43.019	<0.001	0.081		

Discussion

Recently, a lot of emphasis has been given to personalized prophylaxis, which is aimed at managing the genetic risk factor. New knowledge about genetic risk factors from population-based genetic cohort studies requires the creation of different data analytics software to discover the ratio of genetic risk factors and the influence of environmental factors, as well as the impact of comorbidity. In particular, the developed logistic regression model represents an important method of analysis, which is used to confirm the statistical significance of the chosen genetic predictors of IS development in AF patients. The described model can be used as an additional method of predicting the AF-related IS.

Using the developed logistic regression model of IS prediction, depending on predictor genotypes in AF patients, it is possible to take into account genetic factors to improve the complex of preventive measures for a particular individual, this being very important at a low risk of AF-related embolic complications, according to the CHA2DS2-VASc score. The obtained logistic regression model is recommended as an additional method for assessing the risk of IS in young patients with lone AF. To increase the sensitivity of the developed multiple logistic regression model and obtain a more accurate forecast, it is possible to further include in the model the main clinical risk factors for AF-related IS used in the CHA2DS2-VASc score.

Conflict of interest

The authors declare that they have no competing interests.

References

1. Morozova TE, Andrushchishina TB. [Thromboembolic complications prevention in patients with atrial fibrillation]. Farmateka. 2014;9:81–87. [Article in Russian].
2. Napalkov DA, Sokolova AA. [A rational approach to the choice of anticoagulant therapy for the prevention of strokes in patients with nonvalvular atrial fibrillation]. Sechenovsky Bulletin. 2014;1:55–62. [Article in Russian].
3. Novikova NA, Keiko OI. [Ischemic stroke prevention in patients with nonvalvular atrial fibrillation: the role of dabigatran etexilate]. Doctor.Ru. 2014.4:38–41. [Article in Russian].
4. Panchenko EP, Dobrovolsky AB. [Thromboses in cardiology. Mechanisms of development and treatment options]. Sport i Kultura. 1999:217–243. [Article in Russian].
5. Zotova IV, Isaeva MIu, Vanieva OS,Tsymbalova TE, Alekhin MN. Zateishchikov DA. [The system of hemostasis in patients with atrial fibrillation: markers of left auricular thrombosis]. Kardiologiia. 2008;48(2):36–40. [Article in Russian].
6. Sumarokov AV, Mikhailov AA. Cardiac arrythmias. Moscow: Meditsina; 1986. [In Russian].
7. Sychov DA, Kropacheva ES, Ignat'ev IV, Bulytova IuM, Dobrovol'skiĭ AB, Panchenko EP, Kukes VG. [Pharmacogenetics of indirect anticoagulants: value of genotype for improvement of efficacy and safety of therapy]. Kardiologiia. 2006;46(7):72-7. [Article in Russian].
8. Stewart S, Hart CL, Hole DJ, McMurray JJ. A population-based study of the long-term risks associated with atrial fibrillation: 20-year follow-up of the Renfrew/Paisley study. Am J Med. 2002;113(5):359-64.
9. Diagnosis and treatment of atrial fibrillation. Recommendation of RCS, ASSC and ACS. Rus J Cardiol. 2013;4(3):100. [Article in Russian].
10. Kirchhof P, Benussi S, Kotecha D, Ahlsson A, Atar D, Casadei B, et al.; ESC Scientific Document Group. 2016 ESC Guidelines for the management of atrial fibrillation developed in collaboration with EACTS. Eur Heart J. 2016 Oct 7;37(38):2893-2962. doi: 10.1093/eurheartj/ehw210.
11. Serdechnaya EV. Atrial fibrillation: peculiarities of clinical course and choice of treatment strategy. Abstract of ScD Thesis. Moscow; 2008. [In Russian].
12. Serdyuk IYu. Polymorphism of antithrombin genes in patients with ischemic stroke. Abstract of ScD Thesis. Moscow; 2008. [In Russian].
13. Guilyarov M.Yu. Thromboembolic complications in patients with atrial fibrillation: factors influencing the risk of their development and antithrombotic therapy effectiveness. Abstract of ScD Thesis. Moscow; 2011. [In Russian].
14. Shulman VA, Aksyutina NV, Nikulina SYu, Nazarov BV, Dudkina KV, Maksimov VN, Kozlov VV, Kotlovsky MYu, Sinyapko SF, Platunova IM. [Genetic predictors of cardioembolic stroke in patients with atrial fibrillation]. Rus J Cardiol. 2014;10:29-33. [Article in Russian].
15. Pollard J.H. A Handbook of Numerical and Statistical Techniques: With Examples Mainly from the Life Sciences. Cambridge University Press;1979.
16. Frank MB, Reiner AP, Schwartz SM, Kumar PN, Pearce RM, Arbogast PG, et al. The Kozak sequence polymorphism of platelet glycoprotein Ibalpha and risk of nonfatal myocardial infarction and nonfatal stroke in young women. Blood. 2001 Feb 15;97(4):875-9.
17. Zeng Y, Zhang L, Hu Z, Yang Q, Ma M, Liu B, et al. 807C/T polymorphism of platelet glycoprotein Ia gene is associated with cerebral hemorrhage in a Chinese population. Int J Neurosci. 2016;126(8):729-33. doi: 10.3109/00207454.2015.1067891.
18. Hu X, Wang J, Li Y, Wu J, Qiao S, Xu S, et al. The β-fibrinogen gene 455G/A polymorphism associated with cardioembolic stroke in atrial fibrillation with low CHA2DS2-VaSc score. Sci Rep. 2017;7(1):17517. doi: 10.1038/s41598-017-17537-1.
19. Aksyutina NV, Shulman VA, Nikulina SYu, Nazarov BV, Maksimov VN, Plita EV, Kotlovsky MYu, Vereshchagina TD. [Clinical-Genetic riskometer for the ischemic stroke risk assessment in atrial fibrillation]. Rus J Cardiol. 2015;10:42-45. [Article in Russian].